

Statistica Sinica **23** (2013), 1299-1321
doi:<http://dx.doi.org/10.5705/ss.2012.047>

AN ASYMPTOTIC ANALYSIS OF A CLASS OF DISCRETE NONPARAMETRIC PRIORS

Pierpaolo De Blasi¹, Antonio Lijoi² and Igor Prünster¹

¹*University of Torino and Collegio Carlo Alberto, Italy and*

²*University of Pavia and Collegio Carlo Alberto, Italy*

Abstract: In this paper we analyze the asymptotic behaviour of Gibbs-type priors, that represent a natural generalization of the Dirichlet process. After determining their topological support, we investigate their consistency according to the “what if”, or frequentist, approach, that postulates the existence of a “true” distribution P_0 . We provide a full taxonomy of their limiting behaviours: consistency holds essentially always for discrete P_0 , whereas inconsistency may occur for diffuse P_0 . Such findings are further illustrated by means of three special cases admitting closed form expressions and exhibiting a wide range of asymptotic behaviours. For both Gibbs-type priors and discrete nonparametric priors in general, the possible inconsistency should not be interpreted as evidence against their use *tout court*. It rather represents an indication that they are designed for modeling discrete distributions and evidence against their use in the case of diffuse P_0 .

Key words and phrases: Asymptotics, Bayesian consistency, Bayesian nonparametrics, Gibbs-type priors, foundations, species sampling.

1. Introduction

In this paper we study the posterior consistency of Gibbs-type priors recently introduced in Gneden and Pitman (2005). They identify a large class of *discrete* nonparametric priors, which means they select almost surely (a.s.) discrete distributions, and represent probably the most natural generalization of the Dirichlet process, as is argued in Section 2. Several members of this class of nonparametric priors are widely used in practice, for instance, in the contexts of mixture models (Ishwaran and James (2001, 2003); Lijoi, Mena and Prünster (2007c)), linguistics and information retrieval in document modeling (Teh (2006); Teh and Jordan (2010)), species sampling (Lijoi, Mena and Prünster (2007a,b); Navarrete, Quintana and Müller (2008)) and, implicitly, in the context of exchangeable product partition models (Hartigan (1990); Quintana and Iglesias (2003)).

A simple way to introduce Gibbs-type priors is through the system of predictive distributions they induce. To this end, we first lay out the basic framework.

Let $(X_n)_{n \geq 1}$ be an (ideally) infinite sequence of observations, with each X_i taking values in a complete and separable metric space \mathbb{X} . Let $\mathbf{P}_{\mathbb{X}}$ be the set of all probability measures on \mathbb{X} endowed with the topology of weak convergence. In the most commonly employed Bayesian models, $(X_n)_{n \geq 1}$ is assumed to be *exchangeable*, so that, for some Q on $\mathbf{P}_{\mathbb{X}}$,

$$X_i | \tilde{p} \stackrel{\text{i.i.d.}}{\sim} \tilde{p}, \quad \tilde{p} \sim Q. \quad (1.1)$$

Hence, \tilde{p} is a random probability measure on \mathbb{X} whose probability distribution Q is termed *de Finetti measure* and acts as a prior for Bayesian inference. When Q degenerates on a finite dimensional subspace of $\mathbf{P}_{\mathbb{X}}$, the inferential problem is called *parametric*. On the other hand, when the support of Q is infinite-dimensional, one speaks of a *nonparametric* inferential problem and it is generally agreed (Ferguson (1974)) that having a large topological support is a desirable property for a nonparametric prior: we come back to this point later in Section 2. Given a sample (X_1, \dots, X_n) , the predictive distribution coincides with the posterior expected value of \tilde{p} , that is,

$$P(X_{n+1} \in \cdot | X_1, \dots, X_n) = \int_{\mathbf{P}_{\mathbb{X}}} p(\cdot) Q(dp | X_1, \dots, X_n). \quad (1.2)$$

We will deal with discrete priors Q , which implies that a sample (X_1, \dots, X_n) features ties with positive probability: X_1^*, \dots, X_k^* denote the $k \leq n$ distinct observations and n_1, \dots, n_k their frequencies for which $\sum_{i=1}^k n_i = n$. Gibbs-type priors are characterized by predictive distributions (1.2) of the form

$$P(X_{n+1} \in \cdot | X_1, \dots, X_n) = \frac{V_{n+1,k+1}}{V_{n,k}} P^*(\cdot) + \frac{V_{n+1,k}}{V_{n,k}} \sum_{i=1}^k (n_i - \sigma) \delta_{X_i^*}(\cdot), \quad (1.3)$$

where $\sigma \in (-\infty, 1)$, $P^*(dx) := E[\tilde{p}(dx)]$ is a diffuse probability measure representing the prior guess at the shape of \tilde{p} and $\{V_{n,k} : k = 1, \dots, n; n \geq 1\}$ is a set of non-negative weights satisfying the recursion

$$V_{n,k} = (n - \sigma k) V_{n+1,k} + V_{n+1,k+1}. \quad (1.4)$$

Therefore, Gibbs-type priors are characterized by predictive distributions that are a linear combination of the prior guess and a weighted version of the empirical measure. The most widely known prior within this class is the Dirichlet process (Ferguson (1974)).

In this paper we focus on the asymptotic behaviour of Gibbs-type priors and, in particular, investigate posterior consistency according to the “what if” approach of Diaconis and Freedman (1986). Such an approach consists in

assuming that the data $(X_n)_{n \geq 1}$ are independent and identically distributed from some “true” $P_0 \in \mathbf{P}_X$ and in verifying whether the posterior distribution $Q(\cdot | X_1, \dots, X_n)$ accumulates in any neighborhood of P_0 , under a suitable topology. Since Gibbs-type priors are defined on \mathbf{P}_X and are discrete, the appropriate notion of convergence is convergence in the weak topology. Therefore, we aim at establishing whether $Q(A_\epsilon | X_1, \dots, X_n) \rightarrow 1$, a.s.- P_0^∞ , as $n \rightarrow \infty$ and for any $\epsilon > 0$, where A_ϵ denotes a weak neighborhood of P_0 of radius ϵ and P_0^∞ is the infinite product measure $P_0 \times P_0 \times \dots$. In pursuing this plan we first show that “genuinely nonparametric” Gibbs-type priors (a notion that will be clarified in Section 2) have full weak support. We then prove a general structural result on Gibbs-type priors showing that the posterior distribution converges to a point mass at the limiting predictive distribution

$$\alpha P^* + (1 - \alpha)P_0 \quad \alpha \in [0, 1] \quad (1.5)$$

that is a linear combination of the prior guess P^* and the “true” distribution P_0 . This points out that Gibbs-type priors are well-behaved in the limit in the sense of convergence taking place rather than implying consistency. As for the latter to happen, one needs $\alpha = 0$ in (1.5), a feature clearly satisfied in the Dirichlet case. Since a few particular cases of Gibbs-type priors with $\sigma \in (0, 1)$ have already been considered in Jang, Lee, and Lee (2010) and James (2008), attention is focused on the case of $\sigma \in (-\infty, 0)$ for which nothing is known to date and which yield competitive estimators for species estimation in Ecology (Favaro et al. (2012)). A full taxonomy of the asymptotic behaviours is provided. In fact, in deriving the results it is fundamental to distinguish the cases of P_0 discrete and diffuse: in the former case one essentially always has consistency, whereas in the latter we provide a sufficient condition for consistency that has the merit of being close to necessary. This is shown by exhibiting specific priors, which, by a minimal violation of the sufficient condition, already lead to inconsistency. We provide explicit priors exhibiting the extreme limit behaviours $\alpha = 0$ and $\alpha = 1$. The latter corresponds to the worst case scenario where the posterior tends to concentrate around the prior guess P^* and no learning takes place: we refer to such a pathological situation as “total” inconsistency. A third specific prior yields all $\alpha \in (0, 1)$ and serves as interpretation of the two extreme cases.

Our results serve two purposes. The first is to provide a comprehensive analysis of consistency properties of a large and intuitive class of nonparametric priors. This fills in a gap in the current rapidly growing literature on asymptotic properties of Bayesian nonparametric procedures, see Ghosal (2010) for a recent review. We are also concerned with general foundational and methodological questions. Our asymptotic results highlight the fact that discrete nonparametric

priors are actually designed to model discrete distributions and are not appropriate for data coming from diffuse distributions. The typical full weak support property of discrete nonparametric priors led to the thinking that they were suitable models also for diffuse distributions. Consequently, the famous example of inconsistency due to Diaconis and Freedman (1986), involving the use of a Dirichlet process in a semiparametric location problem, was interpreted as an indication of the fact that one needs to be careful with Bayesian nonparametric models in general and, more specifically, with modeling diffuse data with the Dirichlet process. In our opinion, this essentially represented a misunderstanding: its reason probably lies in the fact that the Dirichlet process combines full weak support with consistency for independent and identically distributed data generated from a diffuse P_0 , which is more of a coincidence than a structural property nonparametric priors should possess. We find it wrong to use discrete priors in such contexts and hope to demonstrate this with explicit illustrations. We exhibit a specific nonparametric prior that, in the case of diffuse P_0 , can produce either consistency or “total” inconsistency by simply tuning a scalar parameter. Indeed, consistency is the rule for discrete data generating distributions P_0 or even for diffuse P_0 provided the Gibbs-type prior is used as mixing measure in a hierarchical model; this follows from Ghosal, Ghosh, and Ramamoorthi (1999); Lijoi, Prünster and Walker (2005).

The outline of the paper is as follows. In Section 2, Gibbs-type priors are reviewed and their topological support is investigated. Section 3 contains the general results on the asymptotic behaviour, whereas Section 4 gives specific priors that highlight the various possible asymptotic regimes. Some concluding remarks are provided in Section 5. Online available Supplementary Material provides some technical results, derivations, and proofs.

2. Gibbs-type Priors and Their Topological Support

Modeling data according to a discrete prior Q implies that a sample (X_1, \dots, X_n) has ties with positive probability. Let X_1^*, \dots, X_k^* denote the $k \leq n$ distinct observations and n_1, \dots, n_k their frequencies, $\sum_{i=1}^k n_i = n$. In choosing a specific predictive structure the key quantity to consider is the probability of obtaining a new distinct observation

$$P(X_{n+1} = \text{“new”} \mid X_1, \dots, X_n). \quad (2.1)$$

If Θ is a finite-dimensional parameter entering the specification of \tilde{p} , there are three possibilities for modeling (2.1): (i) $P(X_{n+1} = \text{“new”} \mid X_1, \dots, X_n) = f(n, \Theta)$: the probability of obtaining a new observation depends on the sample

size n but not on the number of distinct observations k and on their frequencies n_1, \dots, n_k ; (ii) $P(X_{n+1} = \text{"new"} | X_1, \dots, X_n) = f(n, k, \Theta)$: dependence is now on both n and k but not on the frequencies n_1, \dots, n_k ; (iii) $P(X_{n+1} = \text{"new"} | X_1, \dots, X_n) = f(n, k, n_1, \dots, n_k, \Theta)$: dependence is on all the sample information. (i) holds if the prior is a Dirichlet process with parameter measure θP^* in which case $P(X_{n+1} = \text{"new"} | X_1, \dots, X_n) = f(n, \Theta) = \theta/(\theta + n)$. Case (ii) corresponds to Gibbs-type priors for which

$$P(X_{n+1} = \text{"new"} | X_1, \dots, X_n) = \frac{V_{n+1,k+1}}{V_{n,k}} \quad (2.2)$$

with the $V_{n,k}$'s satisfying (1.4). In the general situation (iii), serious tractability issues arise: priors have to be studied on a case-by-case basis and typically lead to quite complicated expressions (Favaro, Prünster, and Walker (2011)). See Zabell (1982) and De Blasi et al. (2012) for details on this classification. Thus, the simplifying assumption underlying Gibbs-type priors seems to represent the right compromise between flexibility and tractability. In fact, it is only the probability of obtaining a new observation that does not depend on the frequencies and not the complete prediction rule (1.3). To clarify this point it is useful to interpret (1.3) by means of a two step procedure: X_{n+1} is new with probability $V_{n+1,k+1}/V_{n,k}$; given that X_{n+1} is new, it is sampled independently from P^* and given that X_{n+1} is "old", it coincides with X_i^* with probability $(n_i - \sigma)/(n - k\sigma)$ for $i = 1, \dots, k$, which depends explicitly on n_1, \dots, n_k . When compared to the Dirichlet process, the Gibbs-type framework leads to apparent advantages in species sampling problems (Lijoi, Mena and Prünster (2007a,b)) and also to more robust estimates of the number of components in mixture models (Lijoi, Mena and Prünster (2007c)). As for species sampling, think of having samples of size n with $k' = 1$ and $k'' = n$ distinct species: for the Dirichlet process the probability of observing a new species is $\theta(\theta + n)^{-1}$ in both cases, whereas it explicitly depends on k for other Gibbs-type priors. For instance, if one uses the two-parameter Poisson-Dirichlet process (Pitman (1996)), a Gibbs-type prior, one has the modeling possibilities

$$P(X_{n+1} = \text{"new"} | X_1, \dots, X_n) = \frac{\theta + k\sigma}{\theta + n}, \quad (2.3)$$

where $\sigma \in [0, 1)$ and $\theta > -\sigma$, or $\sigma \in (-\infty, 0)$ and $\theta = x|\sigma|$ for some $x \in \mathbb{N}$. Here, (2.3) is monotonically increasing in k for $\sigma \in (0, 1)$ and monotonically decreasing in k for $\sigma < 0$.

We recall some features of the underlying de Finetti measure Q whose posterior expected value yields the predictive distributions (1.2). Gibbs-type priors

are species sampling models (Pitman (1996)) and can be represented as

$$\tilde{p}(\cdot) = \sum_{i \geq 1} \tilde{p}_i \delta_{Y_i}(\cdot), \quad (2.4)$$

where the weights $(\tilde{p}_n)_{n \geq 1}$ take value on the infinite probability simplex, while the (Y_i) 's are independent and identically distributed from a diffuse P^* , independent of the \tilde{p}_i 's. Clearly, $E[\tilde{p}(\cdot)] = P^*(\cdot)$ which explains the terminology *prior guess* adopted for P^* . Such a framework allows an alternative definition of Gibbs-type priors that coincides with the original one in Gneden and Pitman (2005): Gibbs-type priors are species sampling models (2.4) for which the probability, of obtaining in an n sample k distinct observations with frequencies n_1, \dots, n_k , has the product form

$$V_{n,k} \prod_{i=1}^k (1 - \sigma)_{n_i - 1}, \quad (2.5)$$

for any $n \geq 1$, with $\sigma \in (-\infty, 1)$, the $V_{n,k}$'s satisfying (1.4), and $(a)_m$ denoting the rising factorial $(a)_m = a(a+1) \cdots (a+m-1)$. Such a distribution is known as an exchangeable partition probability function. This concept was introduced by J. Pitman and plays a major role in modern probability theory; see Pitman (2006) and references therein. The special case of the two-parameter Poisson-Dirichlet process has

$$V_{n,k} = \frac{\prod_{i=1}^{k-1} (\theta + i\sigma)}{(\theta + 1)_{n-1}}, \quad (2.6)$$

with $\sigma \in [0, 1)$, $\theta > -\sigma$, or $\sigma \in (-\infty, 0)$, $\theta = x|\sigma|$, $x \in \mathbb{N}$. From (2.6) one obtains (2.3) via (2.2). The two-parameter Poisson-Dirichlet model with $\sigma \in (-\infty, 0)$, $\theta = x|\sigma|$, $x \in \mathbb{N}$ corresponds to an x -variate symmetric Dirichlet distribution with parameter vector $(|\sigma|, \dots, |\sigma|)$.

In Gneden and Pitman (2005) a complete characterization of the underlying de Finetti measure Q is provided and distinguishes three cases according to the value of σ : (i) if $\sigma = 0$, \tilde{p} is either a Dirichlet process or a mixture of Dirichlet processes w.r.t. the total mass parameter θ ; (ii) if $\sigma \in (0, 1)$, then Q is essentially a Poisson-Kingman model based on the stable random measure; see Pitman (2006) and references therein; (iii) if $\sigma < 0$, Q is a mixture of the two-parameter model (2.6) with $\sigma \in (-\infty, 0)$, $\theta = |\sigma|x$, $x \in \mathbb{N}$,

$$V_{n,k} = \sum_{x \geq k} \frac{\prod_{i=1}^{k-1} (x|\sigma| + i\sigma)}{(x|\sigma| + 1)_{n-1}} \pi(x), \quad (2.7)$$

where π is a probability measure on \mathbb{N} and the sum runs over $x \geq k$ since the numerator in the summands corresponding to $x < k$ is 0. Therefore, since in the

case of negative σ the two-parameter model coincides with a x -variate symmetric Dirichlet distribution, one can describe such Gibbs-type priors in terms of the mixture model

$$\begin{aligned} (\tilde{p}_1, \dots, \tilde{p}_k) &\sim \text{Dirichlet}(|\sigma|, \dots, |\sigma|), \\ k &\sim \pi(\cdot). \end{aligned} \quad (2.8)$$

Using the species metaphor (2.7), equivalently (2.8), corresponds to putting a prior π on the number of species k and, conditionally on the number of species being x , these are distributed as a x -variate symmetric Dirichlet distribution. In contrast to the case of $\sigma \geq 0$ where the model assumes the existence of an infinite number of species, the case of $\sigma < 0$ assumes a possibly random but finite number of species. Therefore, in light of the previous considerations, one deduces that if the probability of observing a new species is assumed to depend on n and k but not on n_1, \dots, n_k and moreover the a priori number of species is assumed to be finite (either random or not random), then the model is necessarily (2.8).

Henceforth we restrict attention to “genuinely nonparametric” Gibbs-type priors whose almost sure realizations have support containing a finite number of points that can be equal to any positive integer. These correspond to Gibbs-type priors with either $\sigma \in [0, 1)$ or $\sigma < 0$ such that the support of π in (2.7) is the whole set of positive integers \mathbb{N} . Note that for the “parametric” case of $\sigma < 0$ and π supported by a finite subset of \mathbb{N} , one has consistency for any P_0 in its support by the results of Freedman (1963).

We move on to considering the topological support of Gibbs-type priors. It is widely accepted (Ferguson (1974)) that nonparametric priors should have a large topological support. Since we are dealing with a class of discrete nonparametric priors, this translates to asking Q to have large support in the weak topology. The next result shows that Gibbs-type priors have full weak support, that is their topological support coincides with the space of probability measures whose support is included in the support of the prior guess P^* . In particular, if the support of P^* coincides with \mathbb{X} , the support of Q is the whole space $\mathbf{P}_{\mathbb{X}}$. Such a property is already known in the Dirichlet process case (Ferguson (1973); Majumdar (1992)) and has been recently extended to a class of predictor-dependent nonparametric priors in Barrientos, Jara and Quintana (2012).

Proposition 1. *Let Q be a Gibbs-type prior with prior guess P^* and, in the case $\sigma < 0$, mixing measure π such that $\pi(x) > 0$ for any $x \in \mathbb{N}$. Then the topological support of Q is $\{p \in \mathbf{P}_{\mathbb{X}} : \text{supp}(p) \subset \text{supp}(P^*)\}$.*

The proof is provided in the Supplementary Material. Thus, when used to model the data in species sampling contexts, weak neighborhoods of any given distribution (whose support is included in the support of the prior guess P^*) have

a priori positive probability. This is also a desirable property in the context of mixture models where \tilde{p} acts as a mixing distribution. Indeed, it ensures a high degree of flexibility of the model for any given kernel, and has implications in terms of consistency since one can extend results known for Dirichlet mixtures (Ghosal, Ghosh, and Ramamoorthi (1999); Lijoi, Prünster and Walker (2005)).

3. Posterior Consistency of Gibbs-type Priors

We turn to a study of asymptotic behaviour. We use the notation Q_n for denoting the posterior distribution $Q(\cdot \mid X_1, \dots, X_n)$ of the random probability measure \tilde{p} in (1.1), conditional on the sample X_1, \dots, X_n . Assuming the data are independent and identically distributed from some “true” distribution P_0 in $\mathbf{P}_{\mathbb{X}}$, we are interested in checking whether Q_n concentrates, as n increases, in a weak neighbourhood of some element, say P' , in $\mathbf{P}_{\mathbb{X}}$, almost surely with respect to the infinite product measure P_0^∞ . If A'_ε is a weak neighbourhood of P' with radius $\varepsilon > 0$, we establish conditions under which

$$Q_n(A'_\varepsilon) \rightarrow 1 \quad \text{a.s.-}P_0^\infty \quad (3.1)$$

as $n \rightarrow \infty$ and for any $\varepsilon > 0$. More importantly, we identify cases in which $P' = P_0$, which corresponds to Q being *weakly consistent* in the frequentist sense.

Weak consistency of the Dirichlet process prior is quite straightforward to prove by investigating the asymptotic behaviour of the posterior expected value, the predictive distributions (1.2), and the posterior variance. As the Dirichlet process prior is a special case of Gibbs-type prior, we adopt a similar strategy in this more general framework. Since the predictive distributions in (1.3) characterize Gibbs-type priors, the validity of (3.1) depends on the limiting behaviour of the weights V_{n, κ_n} . We use the notation κ_n to denote the number of blocks in the partition of the first n observations: $\kappa_n := 1 + \sum_{j=2}^n \mathbb{1}_{D_{j-1}}(X_j)$ with $D_{j-1} = \{X_1, \dots, X_{j-1}\}^c$ and $\mathbb{1}_A$ denotes the indicator function of set A . For the asymptotics of κ_n with respect P_0^∞ , different choices of P_0 yield different limiting behaviours. Thus, if P_0 is discrete with N point masses, $N \in \mathbb{N} \cup \{\infty\}$, then $P_0^\infty(\lim_n \kappa_n = N) = 1$ and $P_0^\infty(\lim_n n^{-1} \kappa_n = 0) = 1$ even if $N = \infty$, while, if P_0 is diffuse, $P_0^\infty(\kappa_n = n) = 1$ for any $n \geq 1$. Henceforth we focus on these two cases and adopt the notations $\kappa_n \ll_{a.s.} n$ and $\kappa_n \sim_{a.s.} n$, for $\kappa_n/n \rightarrow 0$ and $\kappa_n/n \rightarrow 1$ a.s.- P_0^∞ , respectively. See Remark 2 for a discussion of the case where P_0 is a combination of a discrete and a diffuse component.

In order to establish the validity of (3.1), for some P' , one needs to investigate the asymptotics for $V_{n+1, \kappa_n+1}/V_{n, \kappa_n}$ under P_0^∞ . Indeed, in what follows we

assume that the probability of recording a new distinct observation at step $n + 1$

$$\frac{V_{n+1, \kappa_{n+1}}}{V_{n, \kappa_n}} \quad \text{converges} \quad \text{a.s.-}P_0^\infty \quad (\text{H})$$

as $n \rightarrow \infty$, and that the limit is identified by some constant $\alpha \in [0, 1]$. For all Gibbs-type priors for which an explicit expression of the V_{n, κ_n} 's is known, (H) holds true regardless of whether P_0 is discrete or diffuse. The role of condition (H) is also transparent: it determines the asymptotics of the predictive distribution and identifies the possible element P' in $\mathbf{P}_{\mathbb{X}}$ for which (3.1) holds true.

Theorem 1. *Let \tilde{p} be a Gibbs-type prior with prior guess $P^* = \mathbb{E}[\tilde{p}]$, whose support coincides with \mathbb{X} , and assume condition (H) holds true. If $(X_i)_{i \geq 1}$ is a sequence of independent and identically distributed random elements from some probability distribution P_0 that is either discrete or diffuse, then the posterior converges weakly, a.s.- P_0^∞ , to a point mass at $\alpha P^*(\cdot) + (1 - \alpha)P_0(\cdot)$.*

According to Theorem 1, weak consistency is guaranteed in the trivial case of $P^* = P_0$, which is excluded henceforth, and when $\alpha = 0$: therefore, it is sufficient to check whether the probability of obtaining a new observation, given previously recorded data, converges to 0, a.s.- P_0^∞ . One can ask whether there are circumstances leading to $\alpha = 1$, which corresponds to the posterior concentrating around the prior guess P^* , a situation we refer to as “total” inconsistency. A specific prior for this is provided in Section 4. Note that Theorem 1 includes, as a special case, Proposition 1 of James (2008) in which \tilde{p} is a two-parameter Poisson-Dirichlet process with parameters (σ, θ) such that $\sigma \in [0, 1]$ and $\theta > -\sigma$. In fact, in the two-parameter Poisson-Dirichlet case, it is immediate to see from (2.3) that when P_0 is discrete, $\kappa_n \ll_{a.s.} n$, we have $\alpha = 0$, implying consistency. When P_0 is diffuse, $\kappa_n \sim_{a.s.} n$, we have $\alpha = \sigma$, hence inconsistency unless $\sigma = 0$. See also Jang, Lee, and Lee (2010, Thm. 1). Let us now provide a proof of the stated result. The key ingredient in the proof of Theorem 1 is an upper bound on the posterior variance $\text{Var}[\tilde{p}(A) \mid X_1, \dots, X_n]$ that is of independent interest. See Remark 1.

Proof of Theorem 1. We show that, under (H), the posterior variance of $\tilde{p}(A)$, given a sample $\mathbf{X}_{\kappa_n}^{(n)} = (X_1, \dots, X_n)$ featuring $\kappa_n \leq n$ distinct values, converges to 0, a.s.- P_0^∞ . To this end, we establish an upper bound for $\text{Var}[\tilde{p}(A) \mid \mathbf{X}_{\kappa_n}^{(n)}]$.

As in Freedman and Diaconis (1983), we consider the class of semi-norms on $\mathbf{P}_{\mathbb{X}}$ defined by $\|P_1 - P_2\|_{\mathcal{A}}^2 = \sum_{i=1}^{\infty} [P_1(A_i) - P_2(A_i)]^2$ for a generating sequence of measurable partitions $\mathcal{A} = \{A_i\}_{i=1}^{\infty}$ of \mathbb{X} . Indeed, convergence under such semi-norms implies weak convergence. Note that $\mathbb{E}[\|\tilde{p} - \mathbb{E}[\tilde{p} \mid \mathbf{X}_{\kappa_n}^{(n)}]\|_{\mathcal{A}}^2 \mid \mathbf{X}_{\kappa_n}^{(n)}] =$

$\sum_{i=1}^{\infty} \text{Var} [\tilde{p}(A_i) | \mathbf{X}_{\kappa_n}^{(n)}]$. Hence, we are going to show that

$$\sum_{i=1}^{\infty} \text{Var} [\tilde{p}(A_i) | \mathbf{X}_{\kappa_n}^{(n)}] \rightarrow 0 \quad \text{a.s.-}P_0^{\infty}$$

as $n \rightarrow \infty$ for any partition \mathcal{A} , which implies that the posterior concentrates in a weak-neighbourhood of the predictive distribution. See also James (2008) for a similar approach in the specific case of the two-parameter Poisson-Dirichlet process. Let $g_{c,d}^{a,b}(n) = V_{n+a,\kappa_n+b}/V_{n+c,\kappa_n+d}$ with a, b, c and d non-negative integers such that $a \geq c$ and $b \geq d$. Exchangeability implies

$$\begin{aligned} E[\tilde{p}(A)^2 | \mathbf{X}_{\kappa_n}^{(n)}] &= \int_A P(X_{n+2} \in A | \mathbf{X}_{\kappa_n}^{(n)}, X_{n+1} = x) P(X_{n+1} \in dx | \mathbf{X}_{\kappa_n}^{(n)}) \\ &= g_{0,0}^{1,1}(n) \int_A P(X_{n+2} \in A | \mathbf{X}_{\kappa_n}^{(n)}, X_{n+1} = x) P^*(dx) \\ &\quad + g_{0,0}^{1,0}(n) \sum_{j=1}^{\kappa_n} \delta_{X_j^*}(A) (n_j - \sigma) P(X_{n+2} \in A | \mathbf{X}_{\kappa_n}^{(n)}, X_{n+1} = X_j^*) \end{aligned}$$

for any $A \in \mathcal{X}$, where $X_1^*, \dots, X_{\kappa_n}^*$ are the κ_n distinct values that partition $\mathbf{X}_{\kappa_n}^{(n)}$ and δ_c is the unit mass concentrated at a point c in \mathbb{X} . After some tedious and lengthy algebra, one gets to

$$\begin{aligned} E[\tilde{p}(A)^2 | \mathbf{X}_{\kappa_n}^{(n)}] &= g_{0,0}^{2,0}(n) \sum_{i,j=1}^k (n_i - \sigma)(n_j + \delta_{i,j} - \sigma) \delta_{X_i^*}(A) \delta_{X_j^*}(A) + 2g_{0,0}^{2,1}(n) \sum_{i=1}^{\kappa_n} (n_i - \sigma) \delta_{X_i^*}(A) P^*(A) \\ &\quad + g_{0,0}^{2,2}(n) P^*(A)^2 + g_{0,0}^{2,1}(n) (1 - \sigma) P^*(A), \end{aligned}$$

where $\delta_{i,j}$ is the Kronecker δ function and note that we have also relied on the diffuseness of P^* . If

$$\tilde{P}_{n,\kappa_n} = \frac{1}{n - \kappa_n \sigma} \sum_{j=1}^{\kappa_n} (n_j - \sigma) \delta_{X_j^*} \quad (3.2)$$

denotes a weighted empirical distribution at the distinct observations, one can

use the above expression for the posterior second moment of $\tilde{p}(A)$ to obtain

$$\begin{aligned} \text{Var} [\tilde{p}(A)|\mathbf{X}_{\kappa_n}^{(n)}] &= \left(g_{0,0}^{2,0}(n) - (g_{0,0}^{1,0}(n))^2 \right) (n - \sigma\kappa_n)^2 \tilde{P}_{n,\kappa_n}(A)^2 \\ &\quad + g_{0,0}^{2,0}(n)(n - \sigma\kappa_n) \tilde{P}_{n,\kappa_n}(A) \\ &\quad + 2 \left(g_{0,0}^{2,1}(n) - g_{0,0}^{1,0}(n)g_{0,0}^{1,1}(n) \right) (n - \sigma\kappa_n) \tilde{P}_{n,\kappa_n}(A) P^*(A) \\ &\quad + \left(g_{0,0}^{2,2}(n) - (g_{0,0}^{1,1}(n))^2 \right) P^*(A)^2 + g_{0,0}^{2,1}(n)(1 - \sigma) P^*(A). \end{aligned}$$

This can be re-expressed in a more convenient form in terms of

$$I(n, \kappa_n) := 1 - \frac{V_{n+2, \kappa_n+1}}{V_{n+1, \kappa_n+1}} \frac{V_{n, \kappa_n}}{V_{n+1, \kappa_n}}, \quad (3.3)$$

thus yielding $\text{Var} [\tilde{p}(A)|\mathbf{X}_{\kappa_n}^{(n)}] = -I(n, \kappa_n) \left(\mathbb{E}[\tilde{p}(A)|\mathbf{X}_{\kappa_n}^{(n)}] \right)^2 + W_{n, \kappa_n}(A)$ where, using the identities (A1) and (A2) of Lemma 1 in the Supplementary Material,

$$\begin{aligned} W_{n, \kappa_n}(A) &= g_{0,0}^{2,1}(n)(n - \sigma\kappa_n) \tilde{P}_{n, \kappa_n}(A) \\ &\quad \times \left[(g_{1,0}^{2,0}(n) - g_{1,1}^{2,1}(n))(n - \sigma\kappa_n) \tilde{P}_{n, \kappa_n}(A) + g_{1,0}^{2,0}(n) \right] \\ &\quad + g_{0,0}^{1,1}(n) P^*(A) \left[(g_{1,1}^{2,2}(n) - g_{1,0}^{2,1}(n)) P^*(A) + g_{1,1}^{2,1}(n)(1 - \sigma) \right] \\ &= I(n, \kappa_n) \mathbb{E}[\tilde{p}(A)|\mathbf{X}_{\kappa_n}^{(n)}] + g_{0,0}^{1,1}(n)(g_{1,1}^{2,2}(n) - g_{1,0}^{2,1}(n)) P^*(A) [P^*(A) - 1] \\ &\quad + g_{0,0}^{1,0}(n) \left(g_{1,0}^{2,0}(n) - g_{1,0}^{2,1}(n) \right) (n - \sigma\kappa_n)^2 \tilde{P}_{n, \kappa_n}(A) [\tilde{P}_{n, \kappa_n}(A) - 1]. \end{aligned}$$

Since $(\tilde{P}_{n, \kappa_n}(A) \vee P^*(A)) \leq 1$,

$$\text{Var} [\tilde{p}(A)|\mathbf{X}_{\kappa_n}^{(n)}] \leq I(n, \kappa_n) \mathbb{E}[\tilde{p}(A)|\mathbf{X}_{\kappa_n}^{(n)}] \left(1 - \mathbb{E}[\tilde{p}(A)|\mathbf{X}_{\kappa_n}^{(n)}] \right) + Z_{n, \kappa_n}(A),$$

where

$$\begin{aligned} Z_{n, \kappa_n}(A) &= g_{0,0}^{1,0}(n)(n - \sigma\kappa_n)^2 \tilde{P}_{n, \kappa_n}(A) (g_{1,1}^{2,1}(n) - g_{1,0}^{2,0}(n))_+ \\ &\quad + g_{0,0}^{1,1}(n) P^*(A) (g_{1,0}^{2,1}(n) - g_{1,1}^{2,2}(n))_+ \end{aligned} \quad (3.4)$$

and, for any a in \mathbb{R} , $a_+ := \max\{a, 0\}$. Use again (A1) and (A2) of Lemma 1 to get

$$\begin{aligned} Z_{n, \kappa_n}(A) &= g_{0,0}^{1,0}(n)(n - \sigma\kappa_n) \tilde{P}_{n, \kappa_n}(A) (g_{1,0}^{2,0}(n) - I(n, \kappa_n))_+ \\ &\quad + g_{0,0}^{1,1}(n) P^*(A) (g_{1,1}^{2,1}(n)(1 - \sigma) - I(n, \kappa_n))_+. \end{aligned}$$

Set now, for any $a \in \mathbb{R}$, $a_- := a - a_+$ and define

$$J(n, \kappa_n) := \left(\frac{V_{n+2, \kappa_n+1}}{V_{n+1, \kappa_n+1}} (1 - \sigma_-) - I(n, \kappa_n) \right)_+ . \quad (3.5)$$

One notes that $(g_{1,1}^{2,1}(n)(1 - \sigma) - I(n, \kappa_n))_+ \leq J(n, \kappa_n)$, and

$$(g_{1,0}^{2,0}(n) - I(n, \kappa_n))_+ \leq (g_{1,1}^{2,1}(n) - I(n, \kappa_n))_+ \leq J(n, \kappa_n).$$

This implies that $Z_{n, \kappa_n}(A) \leq J(n, \kappa_n) \mathbb{E}[\tilde{p}(A) | \mathbf{X}_{\kappa_n}^{(n)}]$, which in turn yields

$$\begin{aligned} \text{Var} [\tilde{p}(A) | \mathbf{X}_{\kappa_n}^{(n)}] &\leq I(n, \kappa_n) \mathbb{E}[\tilde{p}(A) | \mathbf{X}_{\kappa_n}^{(n)}] \left(1 - \mathbb{E}[\tilde{p}(A) | \mathbf{X}_{\kappa_n}^{(n)}] \right) \\ &\quad + J(n, \kappa_n) \mathbb{E}[\tilde{p}(A) | \mathbf{X}_{\kappa_n}^{(n)}] \end{aligned} \quad (3.6)$$

for any A in \mathcal{X} . The upper bound (3.6), combined with $x(1 - x) \leq 1$ for any $x \in [0, 1]$, leads to

$$\sum_{i=1}^{\infty} \text{Var} [\tilde{p}(A_i) | \mathbf{X}_{\kappa_n}^{(n)}] \leq I(n, \kappa_n) + J(n, \kappa_n).$$

We need to show that $J(n, \kappa_n) + I(n, \kappa_n) \rightarrow 0$ a.s.- P_0^∞ as $n \rightarrow \infty$. In the sequel we omit the a.s.- P_0^∞ specification and explicitly use it when possible confusion may arise. By virtue of condition (H), with the limit identified by a value α in $[0, 1]$, one has $(V_{n+1, \kappa_n}/V_{n, \kappa_n})(n - \sigma\kappa_n) \rightarrow (1 - \alpha)$. Hence

$$1 - I(n, \kappa_n) = \frac{V_{n+2, \kappa_n+1}/V_{n+1, \kappa_n+1}}{V_{n+1, \kappa_n}/V_{n, \kappa_n}} \sim \frac{n - \kappa_n\sigma}{n + 1 - (\kappa_n + 1)\sigma}$$

and one can conclude that $I(n, \kappa_n) \rightarrow 0$, as $n \rightarrow \infty$. It follows also that $J(n, k) \rightarrow 0$ as long as $(1 - \sigma_-)V_{n+2, \kappa_n+1}/V_{n+1, \kappa_n+1} \rightarrow 0$, but the latter is also implied by condition (H) since $V_{n+2, \kappa_n+1}/V_{n+1, \kappa_n+1} \sim (1 - \alpha)/(n + 1 - \sigma(\kappa_n + 1))$. The proof is completed after noting that, if P_0 is either discrete or diffuse, the weighted empirical distribution \tilde{P}_{n, κ_n} in (3.2) converges uniformly to P_0 as $n \rightarrow \infty$, a.s.- P_0^∞ , as it can be shown by a suitable adaptation of Glivenko-Cantelli's theorem.

Remark 1. The upper bound for the posterior variance (3.6) is crucial for the determination of the asymptotic behaviour of the posterior distribution and it sheds some light on a distributional property of \tilde{p} that is of independent interest. Its usefulness is also motivated by the fact that the exact expression of posterior variances is typically involved. See, e.g., Jang, Lee, and Lee (2010) for species sampling models and James, Lijoi, and Prünster (2006) for normalized random measures with independent increments. The bound can be simplified under some

further assumptions. Indeed, a close inspection of the arguments used in the proof of Theorem 1 suggests that

$$\text{Var} [\tilde{p}(A) | \mathbf{X}_{\kappa_n}^{(n)}] \leq I(n, \kappa_n) \mathbb{E}[\tilde{p}(A) | \mathbf{X}_{\kappa_n}^{(n)}] \left(1 - \mathbb{E}[\tilde{p}(A) | \mathbf{X}_{\kappa_n}^{(n)}]\right), \quad (3.7)$$

whenever one of the following two inequalities is satisfied:

$$\frac{V_{n+2, \kappa_n}}{V_{n+1, \kappa_n}} - \frac{V_{n+2, \kappa_n+1}}{V_{n+1, \kappa_n+1}} \geq 0, \quad (3.8)$$

$$\frac{V_{n+2, \kappa_n+2}}{V_{n+1, \kappa_n+1}} - \frac{V_{n+2, \kappa_n+1}}{V_{n+1, \kappa_n}} \geq 0. \quad (3.9)$$

Specifically, (3.8) implies (3.9) when $\sigma \in [0, 1)$, and (3.9) implies (3.8) when $\sigma < 0$ as implied by inequality (A3) of Lemma 1 in the Supplementary Material. Since $\text{Var} [\tilde{p}(A) | \mathbf{X}_{\kappa_n}^{(n)}] \leq \mathbb{E}[\tilde{p}(A) | \mathbf{X}_{\kappa_n}^{(n)}] (1 - \mathbb{E}[\tilde{p}(A) | \mathbf{X}_{\kappa_n}^{(n)}])$, for any $n \geq 1$ and A in \mathcal{X} , the validity of one of (3.8)–(3.9) implies that a sharper bound is obtained with the addition of the multiplicative factor $I(n, \kappa_n)$. Such a simplification indeed occurs for the two most widely used instances of Gibbs-type priors. For example, when \tilde{p} is a Dirichlet process with baseline measure θP^* , then $I(n, \kappa_n) = 1/(\theta + n + 1)$ and

$$\text{Var} [\tilde{p}(A) | \mathbf{X}_{\kappa_n}^{(n)}] = \frac{1}{\theta + n + 1} \mathbb{E}[\tilde{p}(A) | \mathbf{X}_{\kappa_n}^{(n)}] \left(1 - \mathbb{E}[\tilde{p}(A) | \mathbf{X}_{\kappa_n}^{(n)}]\right) \quad (3.10)$$

For the two-parameter Poisson-Dirichlet process model with $\theta > 0$ and $\sigma \in (0, 1)$, we recover the bound given in James (2008) as a special case of our general result. Indeed, one can easily check that (3.8) is valid, $I(n, \kappa_n) = 1/(\theta + n + 1)$ and (3.10) holds true with equality replaced by strict inequality.

Remark 2. The case of P_0 neither fully discrete nor diffuse is similar to the case of diffuse P_0 since the diffuse part determines the frequentist asymptotic behaviour by setting the pace of κ_n . This case has been considered in some detail in Jang, Lee, and Lee (2010). Let $P_0 = \lambda P_0^{(d)} + (1 - \lambda) \sum_{j \geq 1} p_{0,j} \delta_{z_j}$, for any $\lambda \in (0, 1)$, with $P_0^{(d)}$ being a diffuse probability measure on \mathbb{X} , $z_j \in \mathbb{X}$, $p_{0,j} \geq 0$ for any $j \geq 1$, and $\sum_j p_{0,j} = 1$. Since in such a case $\kappa_n/n \rightarrow \lambda$, P_0^∞ -a.s., it can be shown that the weighted empirical distribution \tilde{P}_{n, κ_n} defined in (3.2) converges uniformly to $\rho P_0^{(d)} + (1 - \rho) \sum_{j \geq 1} p_{0,j} \delta_{z_j}$, where $\rho = (\lambda - \lambda\sigma)/(1 - \lambda\sigma)$ as $n \rightarrow \infty$. According to the arguments used in the proof of Theorem 1, the posterior converges weakly to a point mass at

$$\alpha P^* + (1 - \alpha) \left(\frac{\lambda - \lambda\sigma}{1 - \lambda\sigma} P_0^{(d)} + \frac{1 - \lambda}{1 - \lambda\sigma} \sum_{j \geq 1} p_{0,j} \delta_{z_j} \right).$$

Excluding, as before, the trivial case $P^* = P_0$ or one of its components, consistency is then achieved if α converges to 0 and, at the same time, ρ coincides with λ . For the latter to happen one either needs $\lambda \in \{0, 1\}$, the purely diffuse and purely discrete cases, or $\sigma = 0$.

In order to complete the picture one needs to identify those situations in which $\alpha = 0$ so that $P' = P_0$ and weak consistency is achieved. For the case $\sigma \in (0, 1)$, some results for the special instances of Gibbs-type priors admitting closed form predictive structure have been derived in James (2008) and in Jang, Lee, and Lee (2010). In contrast, for the case $\sigma < 0$, no results are available in the literature and we focus attention on this subclass of Gibbs-type priors. Theorem 2 gives sufficient conditions for consistency in terms of the tail behaviour of the mixing distribution π on the positive integers \mathbb{N} in (2.8).

Theorem 2. *Let \tilde{p} be a Gibbs-type prior with parameter $\sigma < 0$, mixing measure π and prior guess P^* whose support coincides with \mathbb{X} . Then the posterior is consistent*

(i) *at any discrete P_0 if for sufficiently large x*

$$\frac{\pi(x+1)}{\pi(x)} \leq 1; \quad (\text{T1})$$

(ii) *at any diffuse P_0 if for sufficiently large x and for some $M < \infty$*

$$\frac{\pi(x+1)}{\pi(x)} \leq \frac{M}{x}. \quad (\text{T2})$$

It is worth noting a few implications of (T1) and (T2). Condition (T1) is an extremely mild assumption on the regularity of the tail of the mixing π , it requires $x \mapsto \pi(x)$ to be ultimately decreasing, a condition met by the commonly used probability measures on \mathbb{N} . Nonetheless, one could construct *ad hoc* examples where such a condition fails to be true. For instance, the mixture

$$\pi(x) = a(1-p_1)p_1^{x-1} \mathbb{1}_{\cup_k \{2k\}}(x) + (1-a)(1-p_2)p_2^{x-1} \mathbb{1}_{\cup_k \{2k+1\}}(x)$$

for some a, p_1 and p_2 in $(0, 1)$, does not satisfy (T1). This is not a sign of inconsistency but rather of presumably consistent cases not covered by the sufficient condition (T1). On the other hand, condition (T2) requires the tail of π to be sufficiently light. This is indeed a binding condition and is close to being necessary. This will become clear when we deal with some specific priors in Section 4. As a matter of fact, we describe situations ranging from weak consistency, where (T2) holds true, to inconsistency and “total” inconsistency according as to the

heaviness of the tails of the mixing distribution π that is chosen. The heavier the tails, the further from P_0 is the limiting P' in (3.1).

Proof of Theorem 2. The proof amounts to showing that (H) holds true with $\alpha = 0$ so that consistency follows by Theorem 1. Let $V_{n,\kappa_n} = \sum_{x \geq \kappa_n} V_{n,\kappa_n}^{\sigma,x} \pi(x)$ where

$$V_{n,\kappa_n}^{\sigma,x} = \frac{|\sigma|^{k-1} \prod_{i=1}^{k-1} (x-i)}{(x|\sigma| - 1)_{n+1}}.$$

Then $V_{n,\kappa_n} = \sum_{y \geq 0} v_{n,\kappa_n}(y)$ where

$$v_{n,\kappa_n}(y) = \frac{|\sigma|^{\kappa_n-n} (y+1)_{\kappa_n-1}}{(\kappa_n + y + 1/|\sigma|) \cdots (\kappa_n + y + [(n-1)/|\sigma|])} \pi(y + \kappa_n).$$

After some algebra,

$$V_{n+1,\kappa_{n+1}} = \sum_{y \geq 0} x_{n,\kappa_n}(y) \frac{\pi(\kappa_n + y + 1)}{\pi(\kappa_n + y)} v_{n,\kappa_n}(y), \quad (3.11)$$

where $x_{n,\kappa_n}(y) = (\kappa_n + y) a_{n,\kappa_n}(y) / (n/|\sigma| + \kappa_n + y + 1)$ and

$$a_{n,\kappa_n}(y) = \prod_{i=1}^{n-1} \frac{(\kappa_n + y + i/|\sigma|)}{(\kappa_n + y + 1 + i/|\sigma|)}.$$

We start by considering discrete P_0 . This yields $\kappa_n \ll_{a.s.} n$ and we assume $P_0^\infty[\lim_n \kappa_n = \infty] = 1$; indeed when $P_0^\infty[\lim_n \kappa_n < \infty] = 1$ the proof of Theorem 2(i) is straightforward. For n large enough

$$\begin{aligned} \frac{V_{n+1,\kappa_{n+1}}}{V_{n,\kappa_n}} &\leq \frac{1}{V_{n,\kappa_n}} \sum_{y \geq 0} x_{n,\kappa_n}(y) v_{n,\kappa_n}(y) \\ &\leq x_{n,\kappa_n}(\kappa_n) \frac{\sum_{y=0}^{\kappa_n} v_{n,\kappa_n}(y)}{V_{n,\kappa_n}} + \sum_{y \geq \kappa_n+1} x_{n,\kappa_n}(y) \frac{v_{n,\kappa_n}(y)}{V_{n,\kappa_n}} \\ &\leq x_{n,\kappa_n}(\kappa_n) + \frac{1}{V_{n,\kappa_n}} \sum_{y \geq \kappa_n+1} v_{n,\kappa_n}(y), \end{aligned} \quad (3.12)$$

where we used (T1) in the first inequality, the monotonicity of $y \mapsto x_{n,\kappa_n}(y)$ in the second inequality and $x_{n,\kappa_n}(y) \leq 1$ in the last inequality. Note that, as $n \rightarrow \infty$,

$$x_{n,\kappa_n}(\kappa_n) = \frac{2\kappa_n}{(n/|\sigma|) + 2\kappa_n + 1} a_{n,\kappa_n}(\kappa_n) \rightarrow 0 \quad (3.13)$$

a.s.- P_0^∞ , since $a_{n,\kappa_n}(y) \leq 1$ for any y and n . As for the second summand in (3.12), note that

$$\begin{aligned} \frac{1}{V_{n,\kappa_n}} \sum_{y \geq \kappa_n+1} v_{n,\kappa_n}(y) &= \frac{v_{n,\kappa_n}(\kappa_n)}{V_{n,\kappa_n}} \sum_{y \geq 0} \frac{v_{n,\kappa_n}(\kappa_n + y + 1)}{v_{n,\kappa_n}(\kappa_n)} \\ &\leq \frac{v_{n,\kappa_n}(\kappa_n)}{v_{n,\kappa_n}(\kappa_n - 1)} \sum_{y \geq 0} \prod_{j=0}^y \frac{v_{n,\kappa_n}(\kappa_n + j + 1)}{v_{n,\kappa_n}(\kappa_n + j)}. \end{aligned}$$

By virtue of

$$\frac{v_{n,\kappa_n}(y+1)}{v_{n,\kappa_n}(y)} = \frac{\kappa_n + y}{y+1} a_{n,\kappa_n}(y) \frac{\pi(\kappa_n + y + 1)}{\pi(\kappa_n + y)}$$

and (T1), for n large enough one has

$$\frac{1}{V_{n,\kappa_n}} \sum_{y \geq \kappa_n+1} v_{n,\kappa_n}(y) \leq 2a_{n,\kappa_n}(\kappa_n - 1) \sum_{y \geq 0} \prod_{j=0}^y \frac{2\kappa_n + j}{\kappa_n + j + 1} a_{n,\kappa_n}(\kappa_n + j).$$

In view of (3.12) and (3.13) one just needs to prove that

$$a_{n,\kappa_n}(\kappa_n - 1) \rightarrow 0 \quad (3.14)$$

as $n \rightarrow \infty$, and

$$\sum_{y \geq 0} \prod_{j=0}^y \frac{2\kappa_n + j}{\kappa_n + j + 1} a_{n,\kappa_n}(\kappa_n + j) < \infty \quad (3.15)$$

for sufficiently large n . To this aim, note that

$$a_{n,\kappa_n}(\kappa_n - 1) = \prod_{j=1}^{n-1} \left(1 - \frac{|\sigma|}{2\kappa_n|\sigma| + j} \right).$$

If $S_{n,k} = |\sigma| \sum_{j=1}^{n-1} (k|\sigma| + j)^{-1}$, using the inequalities $(1 - R)^{x/R} \leq 1 - x \leq e^{-x}$ for any $0 \leq x \leq R \leq 1$, it easily follows that

$$\left(1 - \frac{1}{2\kappa_n} \right)^{2\kappa_n S_{n,2\kappa_n}} \leq a_{n,\kappa_n}(\kappa_n - 1) \leq e^{-S_{n,2\kappa_n}}. \quad (3.16)$$

Moreover, $(1 - 1/(2\kappa_n))^{2\kappa_n} \rightarrow e^{-1}$ and, as $n \rightarrow \infty$,

$$S_{n,2\kappa_n} \sim \log \left(\frac{n + 2\kappa_n|\sigma| - 1}{2\kappa_n|\sigma|} \right)^{|\sigma|}. \quad (3.17)$$

These, combined with (3.16), lead to the asymptotic evaluation

$$a_{n,\kappa_n}(\kappa_n - 1) \sim \left(\frac{|\sigma|2\kappa_n}{n + |\sigma|2\kappa_n - 1} \right)^{|\sigma|} \quad (3.18)$$

as $n \rightarrow \infty$. As for (3.15), the y -th term of the series can be written as $\varphi_n(0) \cdots \varphi_n(y)$ where

$$\varphi_n(y) = \frac{2\kappa_n + y}{\kappa_n + y + 1} a_{n,\kappa_n}(\kappa_n + y) = \frac{2\kappa_n + y}{\kappa_n + y + 1} \prod_{j=1}^{n-1} \left(1 - \frac{|\sigma|}{(2\kappa_n + y + 1)|\sigma| + j} \right).$$

Adapting the arguments used in (3.16) and (3.17), it can be shown that

$$\varphi_n(y) \sim \frac{2\kappa_n + y}{\kappa_n + y + 1} \left(\frac{|\sigma|(2\kappa_n + y + 1)}{|\sigma|(2\kappa_n + y + 1) + n - 1} \right)^{|\sigma|}$$

as $n \rightarrow \infty$, cfr. (3.18). Next, for $y \rightarrow \infty$, a first order Taylor expansion yields

$$\begin{aligned} \varphi_n(y) &\sim \left(1 + \frac{\kappa_n - 1}{\kappa_n + y + 1} \right) \left(1 - \frac{n - 1}{|\sigma|(2\kappa_n + y + 1) + n - 1} \right)^{|\sigma|} \\ &= \left(1 + \frac{\kappa_n - 1}{\kappa_n + y + 1} \right) \left(1 - \frac{|\sigma|(n - 1)}{|\sigma|(2\kappa_n + y + 1) + n - 1} \right) + O(y^{-2}) \\ &= 1 + \left(\frac{\kappa_n - 1}{\kappa_n + y + 1} - \frac{|\sigma|(n - 1)}{|\sigma|(2\kappa_n + y + 1) + n - 1} \right) + O(y^{-2}) \\ &= 1 - \frac{n - \kappa_n}{y} + O(y^{-2}). \end{aligned}$$

Finally, the series in (3.15) is convergent since $n - \kappa_n > 0$ (Pólya and Szegő (1978)). This completes the proof of (i).

For diffuse P_0 is diffuse, $\kappa_n = n$ a.s.- P_0^∞ and

$$\begin{aligned} \frac{V_{n+1,n+1}}{V_{n,n}} &\leq \frac{1}{V_{n,n}} \sum_{y \geq 0} x_{n,n}(y) \frac{M}{n + y} v_{n,n}(y) \\ &\leq \frac{1}{V_{n,n}} \frac{M}{n/|\sigma| + n + 1} \sum_{y \geq 0} a_{n,n}(y) v_{n,n}(y) \\ &\leq \frac{1}{V_{n,n}} \frac{M}{n/|\sigma| + n + 1} \sum_{y \geq 0} v_{n,n}(y) = \frac{M}{n/|\sigma| + n + 1}, \end{aligned}$$

where we used (T2) for $x = n + y$ in the first inequality, $n/|\sigma| + n + y + 1 > n/|\sigma| + n + 1$ in the second inequality and $a_{n,n}(y) \leq 1$ in the last inequality. Since the last term goes to 0 as $n \rightarrow \infty$, the proof is complete.

4. Illustrations

By Theorem 2 Gibbs-type priors are consistent when P_0 is discrete, condition (T1) being valid for most commonly used mixing measures π . When P_0 is diffuse one needs to closely investigate the tail behaviour of π and check whether (T2)

holds true. One is then naturally led to ask what happens when (T2) is not satisfied: may in such a case consistency fail to occur even if the “true” P_0 is in the weak support of \tilde{p} ?

In this section we consider three Gibbs-type priors with $\sigma = -1$, each characterized by a specific choice of the mixing distribution π . We have consistency at a discrete P_0 by Theorem 2(i) and therefore focus on the case of P_0 diffuse, where different conclusions are reached. According to the heaviness of the tails of π one can move from a situation where α in Theorem 1 is 0, thus yielding consistency, to a situation where α is arbitrarily close or even equal to $\alpha = 1$.

The first prior is characterized by a heavy-tailed mixing distribution π , that does not admit a finite expected value, condition (T2) is not met and $\alpha = 1$ so that the posterior concentrates around the prior guess P^* , “total” inconsistency. The second prior has a mixing π with light tails that satisfy (T2) in Theorem 2, and results in consistent asymptotic behaviour. In the third case α takes values over the whole unit interval $[0, 1]$ according to a parameter that determines the heaviness of the tail of π . The illustration shows that, if the upper bound in (T2) does not hold, consistency is not achieved and we conclude that (T2) is close to being necessary.

4.1. Gnedin’s Gibbs-type prior

We consider the family of Gibbs-type priors with $\sigma = -1$ recently introduced in Gnedin (2010). It is characterized by the mixing distribution

$$\pi(x) = \frac{\gamma(1-\gamma)_{x-1}}{x!} \mathbb{1}_{\{1,2,\dots\}}(x)$$

for some $\gamma \in (0, 1)$. This distribution arises in discrete renewal theory (Feller (1971, Chap. XII) and in connection with the two-parameter Poisson-Dirichlet process (Pitman (2006)). It is characterized by a heavy tail admitting moments of order less than γ . In order to establish consistency one would like to apply Theorem 2. For a discrete P_0 , (T1) clearly holds and weak consistency is achieved. In contrast, for a diffuse P_0 , (T2) is not satisfied: $\pi(x+1)/\pi(x) = (x-\gamma)/(x+1)$ for any positive integer x and is not eventually bounded by M/x for some constant M . We appeal to a direct calculation to determine consistency or inconsistency.

In Gnedin (2010) it is shown that the V_{n,κ_n} ’s admit the simple closed form expression

$$V_{n,\kappa_n} = \frac{(\kappa_n - 1)!(1-\gamma)_{\kappa_n-1}(\gamma)_{n-\kappa_n}}{(n-1)!(1+\gamma)_{n-1}}$$

and, consequently, the weights of the prediction rule simplify to

$$\frac{V_{n+1,\kappa_n+1}}{V_{n,\kappa_n}} = \frac{\kappa_n(\kappa_n - \gamma)}{n(\gamma + n)}. \quad (4.1)$$

From (4.1) it is easy to see that, if P_0 is diffuse, $\kappa_n = n$, condition (H) holds true with $\alpha = 1$. By Theorem 1 it follows that the weak limit coincides with the prior guess P^* whatever the “true” distribution of the data P_0 . In this explicit setup, it is interesting to have a closer look at the structure of the bound on the posterior variance discussed in Remark 1. It is easy to check that (3.9) holds and so the bound (3.10) applies with $I(n, \kappa_n) = (2n + \gamma + 1)/[(n + 1)(\gamma + n + 1)]$, which does not depend on κ_n . Now, since $I(n, \kappa_n) \rightarrow 0$ as $n \rightarrow \infty$, the posterior concentrates, as n increases, at some P' in $\mathbf{P}_{\mathbb{X}}$, in accordance with Theorem 1. Note then that consistency for the case of discrete P_0 , already established by means of Theorem 2(i), can be deduced from (4.1) combined with Theorem 1: if P_0 is discrete then $\kappa_n \ll_{a.s.} n$ and (4.1) converges to $\alpha = 0$ implying convergence to P_0 in Theorem 1.

4.2. Gibbs-type prior with Poisson mixing

The second Gibbs-type prior we consider takes $\sigma = -1$ and a Poisson mixing distribution π restricted to the positive integers, with parameter $\lambda > 0$

$$\pi(x) = \frac{e^{-\lambda}}{1 - e^{-\lambda}} \frac{\lambda^x}{x!} \mathbb{1}_{\{1,2,\dots\}}(x).$$

Such a π has light tails and condition (T2) is satisfied since $\pi(x+1)/\pi(x) = \lambda/(x+1)$. By Theorem 2(ii), the posterior is consistent when P_0 is diffuse and, *a fortiori*, when P_0 is discrete. Given the Gibbs-type prior at issue admits closed form expressions, the same conclusion can be drawn by direct calculation. The V_{n,κ_n} 's can be expressed as

$$V_{n,\kappa_n} = \pi(\kappa_n) V_{n,\kappa_n}^{-1,\kappa_n} {}_1F_1(\kappa_n; \kappa_n + n; \lambda),$$

where ${}_1F_1(a; b; z) = \sum_{j \geq 0} \frac{(a)_j}{j! (b)_j} z^j$ is, for any a, b and z in \mathbb{R} , the confluent hypergeometric function. Therefore, one has that

$$\frac{V_{n+1,\kappa_n+1}}{V_{n,\kappa_n}} = \frac{\lambda \kappa_n}{(n + \kappa_n + 1)(n + \kappa_n)} \frac{{}_1F_1(\kappa_n; \kappa_n + n; \lambda)}{{}_1F_1(\kappa_n + 1; \kappa_n + n + 2; \lambda)}. \quad (4.2)$$

With P_0 diffuse, $\kappa_n = n$ for any n a.s. $-P_0^\infty$ and one has

$$\frac{V_{n+1,n+1}}{V_{n,n}} \sim \frac{\lambda}{2(2n+1)} \rightarrow 0 \quad (4.3)$$

as $n \rightarrow \infty$. Details for the derivation of (4.3) are provided in the Supplementary Material. Hence, for the case of diffuse P_0 , we have shown by direct calculation that the probability of observing a new species converges to $\alpha = 0$, which by Theorem 1 implies consistency. This is clearly in agreement with the conclusion drawn from Theorem 2(ii) by looking at the tails of the mixing distribution π .

4.3. Gibbs-type prior with geometric mixing

Another sub-family of Gibbs-type priors with $\sigma = -1$ is identified by a geometric mixing distribution

$$\pi(x) = (1 - \eta)\eta^{x-1} \mathbb{1}_{\{1,2,\dots\}}(x)$$

for some $\eta \in (0, 1)$. Here $\pi(x+1)/\pi(x) = \eta$ so that (T2) does not hold. One can only apply Part (i) of Theorem 2 to state consistency for the case of discrete P_0 . For the case of diffuse P_0 not covered by Theorem 2, direct calculation gives

$$V_{n,\kappa_n} = \pi(\kappa_n) V_{n,\kappa_n}^{-1,\kappa_n} {}_2F_1(\kappa_n, \kappa_n + 1; \kappa_n + n; \eta),$$

where ${}_2F_1(a, b; c; z) = \sum_{j \geq 0} ((a)_j (b)_j / j! (c)_j) z^j$ for any a, b , and c in \mathbb{R} and for any z such that $|z| < 1$, is the Gauss hypergeometric function. One has

$$\frac{V_{n+1,\kappa_n+1}}{V_{n,\kappa_n}} = \frac{\eta \kappa_n (\kappa_n + 1)}{(n + \kappa_n + 1)(n + \kappa_n)} \frac{{}_2F_1(\kappa_n, \kappa_n + 1; \kappa_n + n; \eta)}{{}_2F_1(\kappa_n + 1, \kappa_n + 2; \kappa_n + n + 2; \eta)}. \quad (4.4)$$

With P_0 diffuse, one can replace κ_n with n in this ratio to find

$$\frac{V_{n+1,n+1}}{V_{n,n}} \rightarrow \alpha = \frac{2 - \eta - 2\sqrt{1 - \eta}}{\eta} \in [0, 1]. \quad (4.5)$$

Details for the derivation of (4.5) are provided in the Supplementary Material. The limit α in (4.5) can be any point in $[0, 1]$ according to the value of η : by Theorem 1 it follows that we can obtain the whole spectrum of weak limits $\alpha P^*(\cdot) + (1 - \alpha)P_0(\cdot)$ ranging from consistency to “total” inconsistency. In particular, α is increasing in η , so the larger η , the heavier the limiting mass assigned to the prior guess. Small values of η identify a situation similar to the one discussed in Section 4.2 since they yield a light-tailed π . Conversely, large values of η give rise to heavy-tailed π . Here a minimal deviation from (T2) can produce inconsistent behaviours, even extreme ones, suggesting again that (T2) is close to being a necessary condition.

5. Concluding Remarks

For the validation of a statistical model, and of the corresponding inferences, consistency plays a major role. Even in a Bayesian framework, an important prerequisite to any inferential procedure is the specification of a prior that is consistent according to the frequentist approach. If \mathbb{X} is finite, P_0 in the weak support of a discrete nonparametric prior \tilde{p} guarantees consistency (Freedman (1963)). When \mathbb{X} is infinite, inconsistent behaviours can appear. To deal with this one can try to identify classes of priors that are consistent whatever the

choice of P_0 or try to identify the data generating mechanisms the various classes of nonparametric priors are designed for and study consistency w.r.t. choices of P_0 that are compatible with such mechanisms. The seminal contribution to the former is due to Freedman (1963) (see also Fabius (1964)), where the author identifies a class of nonparametric priors, the family of “tail-free” priors, that are consistent for any P_0 , discrete or diffuse, in its weak support. Notably, the Dirichlet process and Pólya-tree priors (Ferguson (1974); Lavine (1992)) belong to this class. However, ensuring consistency for any P_0 is not for free. On the one hand, all tail-free priors (with the exception of the Dirichlet process), and the inferential results they yield, depend heavily on the sequence of nested partitions defining them. On the other hand “tail-freeness” appears to be a quite fragile property: as shown in Freedman and Diaconis (1983) and Diaconis and Freedman (1986), inconsistency can already appear when one considers mixtures of the Dirichlet process. Perhaps it is better to establish what kind of inferential issues a prior can address and study consistency for compatible P_0 ’s, and we have taken this path in this paper. Gibbs-type priors are discrete nonparametric priors and therefore consistency has to be investigated w.r.t. discrete P_0 ’s. The answer we provide is positive in the sense that they are (essentially) always consistent w.r.t. discrete P_0 ’s. Given the nature of the phenomenon to be studied, one can establish in advance whether the “true” distribution of the data is discrete or not. When one considers a diffuse data generating P_0 , that does not fit a framework within which Gibbs-type priors are used, inconsistency can arise, even “total” inconsistency. This is not a reason to dismiss Gibbs-type priors. Inconsistent behaviours, combined with consistency in the case of discrete P_0 , should rather be seen as strong general methodological evidence against the use of discrete nonparametric priors for modeling data generated from diffuse distributions, a common practice, for instance, in survival analysis applications.

Supplementary Material

Supplementary Material available online includes (i) the proof of Proposition 1; (ii) an auxiliary technical lemma, that is used in the proof Theorem 1 and provides some useful results on various quantities related to the weights $V_{n,k}$ defining the partition distribution induced by Gibbs-type prior; (iii) technical details on the specific examples considered in Section 4.

Acknowledgement

The authors would like to thank an associate editor and two referees for important comments and remarks. Special thanks are also due to the consulting editor whose suggestions greatly improved the effectiveness of the presentation.

Subhashis Ghosal, Sasha Gnedin and Judith Rousseau are gratefully acknowledged for several insightful discussions. This work was supported by the European Research Council (ERC) through StG “N-BNP” 306406. The paper has been completed while A. Lijoi and I. Prünster were visiting ICERM at Brown University. The hospitality of the Institute is greatly acknowledged.

References

- Barrientos, A.F., Jara, A. and Quintana, F. A. (2012). On the support of MacEachern’s dependent Dirichlet processes and extensions. *Bayesian Anal.* **7**, 277-310.
- De Blasi, P., Favaro, S., Lijoi, A., Mena, R. H., Prünster, I. and Ruggiero, M. (2012). Are Gibbs-type priors the most natural generalization of the Dirichlet process? Technical Report.
- Diaconis, P. and Freedman, D. (1986). On the consistency of Bayes estimates *Ann. Statist.* **14**, 1-26.
- Fabius, J. (1964). Asymptotic behavior of Bayes’ estimates. *Ann. Math. Statist.* **35**, 846-856.
- Favaro, S., Lijoi, A., Mena, R. H. and Prünster, I. (2012). Bayesian nonparametric estimators derived from Gibbs-type priors with finitely many types. In preparation.
- Favaro, S., Prünster, I. and Walker, S.G. (2011). On a class of random probability measures with general predictive structure. *Scand. J. Statist.* **38**, 359-376.
- Feller, W. (1971). *An Introduction to Probability Theory and its Applications*. Vol. 2, 2nd edition. Wiley, New York.
- Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.* **1**, 209-230.
- Ferguson, T. S. (1974). Prior distributions on spaces of probability measures. *Ann. Statist.* **2**, 615-629.
- Freedman, D. (1963). On the asymptotic behavior of Bayes’ estimates in the discrete case. *Ann. Math. Statist.* **34**, 1386-1403.
- Freedman, D. and Diaconis, P. (1983). On inconsistent Bayes estimates in the discrete case. *Ann. Statist.* **11**, 1109-1118.
- Ghosal, S. (2010). The Dirichlet process, related priors, and posterior asymptotics. In *Bayesian Nonparametrics* (Edited by N. Hjort, C. Holmes, P. Müller and S. Walker), 35-79. Cambridge Univ. Press, Cambridge.
- Ghosal, S., Ghosh, J. K. and Ramamoorthi, R. V. (1999). Posterior consistency of Dirichlet mixtures in density estimation *Ann. Statist.* **27**, 143-158.
- Gnedin, A. (2010). A species sampling model with finitely many types. *Elect. Comm. Probab.* **15**, 79-88.
- Gnedin, A. and Pitman, J. (2005). Exchangeable Gibbs partitions and Stirling triangles. *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI)* **325**, 83-102.
- Hartigan, J. A. (1990). Partition models. *Comm. Statist. Theory Methods* **19**, 2745-2756.
- Ishwaran, H. and James, L. F. (2001). Gibbs sampling methods for stick-breaking priors. *J. Amer. Stat. Ass.* **96**, 161-173.
- Ishwaran, H. and James, L. F. (2003). Generalized weighted Chinese restaurant processes for species sampling mixture models. *Statist. Sinica* **13**, 1211-1235.
- James, L. F. (2008). Large sample asymptotics for the two parameter Poisson Dirichlet process. In *Pushing the Limits of Contemporary Statistics*. (Edited by B. Clarke, S. Ghosal), 187-199. IMS, Hayward.

- James, L.F., Lijoi, A. and Prünster, I. (2006). Conjugacy as a distinctive feature of the Dirichlet process. *Scand. J. Statist.* **33**, 105-120.
- Jang, G. H., Lee, J. and Lee, S. (2010). Posterior consistency of species sampling priors. *Statist. Sinica* **20**, 581-593.
- Lavine, M. (1992). Some aspects of Pólya tree distributions for statistical modelling. *Ann. Statist.* **20**, 1222-1235.
- Lijoi, A., Mena, R.H., and Prünster, I. (2007a). Bayesian nonparametric estimation of the probability of discovering a new species *Biometrika*. **94** 769-786.
- Lijoi, A., Mena, R.H., and Prünster, I. (2007b). A Bayesian nonparametric method for prediction in EST analysis. *BMC Bioinformatics* **8**: 339.
- Lijoi, A., Mena, R.H., and Prünster, I. (2007c). Controlling the reinforcement in Bayesian nonparametric mixture models. *J. Roy. Statist. Soc. B* **69**, 715-740.
- Lijoi, A., Prünster, I. and Walker, S. G. (2005). On consistency of nonparametric normal mixtures for Bayesian density estimation. *J. Amer. Statist. Assoc.* **100**, 1292-1296.
- Majumdar, S. (1992). On topological support of Dirichlet prior. *Statist. Probab. Lett.* **15**, 385-388.
- Navarrete, C., Quintana, F. and Müller, P. (2008). Some issues on nonparametric Bayesian modeling using species sampling models. *Stat. Modell.* **41**, 3-21.
- Pitman, J. (1996). Some developments of the Blackwell-MacQueen urn scheme. In *Statistics, Probability and Game Theory* (Edited by T. S. Ferguson, L. S. Shapley and J. B. MacQueen), 245-267. IMS Lecture Notes Monogr. Ser., Vol. **30**, Hayward.
- Pitman, J. (2006). *Combinatorial Stochastic Processes*. Ecole d'Été de Probabilités de Saint-Flour XXXII. Lecture Notes in Math., vol. **1875**. Springer, Berlin.
- Pólya, G. and Szegő, G. (1978). *Problems and Theorems in Analysis. I. Series, Integral Calculus, Theory of Functions*. Springer-Verlag, Berlin-New York.
- Quintana, F. A. and Iglesias, P.L. (2003). Bayesian clustering and product partition models. *J. R. Stat. Soc. B* **65**, 557-574.
- Teh, Y. W. (2006). A Hierarchical Bayesian Language Model based on Pitman-Yor Processes. In *Proceedings of Coling/ACL 2006*, 985-992.
- Teh, Y. W. and Jordan, M. I. (2010). Hierarchical Bayesian nonparametric models with applications. In *Bayesian Nonparametrics* (Hjort, N.L., Holmes, C.C. Müller, P., Walker, S.G. Eds.), 80-136. Cambridge University Press, Cambridge.
- Zabell, S.L. (1982). W. E. Johnson's "sufficientness" postulate. *Ann. Statist.* **10**, 1090-1099.
- Department of Economics and Statistics, University of Torino, Corso Unione Sovietica 218/bis, 10134 Torino, Italy.
E-mail: pierpaolo.deblasi@unito.it
- Department of Economics and Management, University of Pavia, Via San Felice 5, 27100 Pavia, Italy.
E-mail: lijoi@unipv.it
- Department of Economics and Statistics, University of Torino, Corso Unione Sovietica 218/bis, 10134 Torino, Italy.
E-mail: igor.pruenster@unito.it

(Received February 2012; accepted November 2012)